

Epidemic spreading is always possible on regular networks

Charo I. del Genio

Warwick Mathematics Institute
Centre for Complexity Science
Warwick Infectious Disease Epidemiology Research (WIDER) Centre

University of Warwick

Networks and epidemics

For quite some time now, researchers have been using network models to study natural and engineered systems of diverse nature.

The main appeal of the network approach is the simplification of a complex system into a discrete set of elements interacting via pairwise connections.

This paradigm has been very successful in mathematical epidemiology, where the links represent contacts amongst individuals across which diseases can spread.

Regular networks

In ecological or spatially-embedded epidemiological models, a common approach is to use regular networks, where all the nodes have the same number of links k .

Since the links per node are fixed, the total number of links is proportional to the nodes N , and thus regular networks are sparse.

Diseases recently studied using regular networks include the Citrus Tristeza Virus, the foot-and-mouth disease and the bubonic plague.

Regular networks

In ecological or spatially-embedded epidemiological models, a common approach is to use regular networks, where all the nodes have the same number of links k .



Since the links per node are fixed, the total number of links is proportional to the nodes N , and thus regular networks are sparse.

Diseases recently studied using regular networks include the Citrus Tristeza Virus, the foot-and-mouth disease and the bubonic plague.

Regular networks

In ecological or spatially-embedded epidemiological models, a common approach is to use regular networks, where all the nodes have the same number of links k .



Since the links per node are fixed, the total number of links is proportional to the nodes N , and thus regular networks are sparse.



Diseases recently studied using regular networks include the Citrus Tristeza Virus, the foot-and-mouth disease and the bubonic plague.

Regular networks

In ecological or spatially-embedded epidemiological models, a common approach is to use regular networks, where all the nodes have the same number of links k .



Since the links per node are fixed, the total number of links is proportional to the nodes N , and thus regular networks are sparse.



Diseases recently studied using regular networks include the Citrus Tristeza Virus, the foot-and-mouth disease and the bubonic plague.



Clustering

An important feature of networks that influences their structure and the dynamics of the processes they support is the clustering coefficient c .

The clustering coefficient measures the tendency of the nodes to form triangles, and it can be defined mathematically as the ratio of the closed triplets to the total number of triplets.

Notably, many experimental studies have shown how real-world networks often exhibit high values of the clustering coefficient, and specific random graph models have been developed to reproduce the observed results.

Clustering-driven fragmentation

Notwithstanding the recognized importance of clustering, there is no unified theoretical understanding of its influence on the many properties of networks.

For regular networks, a high clustering coefficient will cause fragmentation into several disconnected components, as an increasing number of links is used to form triangles.

This is quite relevant to epidemic modelling, as the largest component in a network corresponds to the largest number of individuals who can contract an infection during the outbreak of a disease.

The received wisdom

We want to study how the size of the largest component in regular networks changes with c . One recurrent method in epidemiology is the use of moment closure techniques.

In our case, moment closure results indicate the existence of a phase transition for a critical value $c^*=(k-2)/(k-1)$.

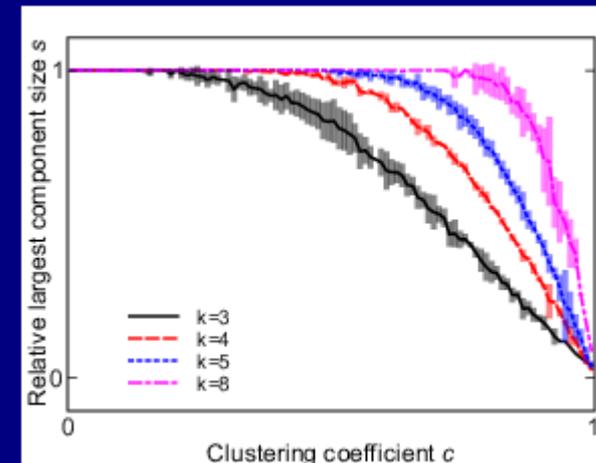
When $c < c^*$, the largest component size grows linearly with N ; instead, when $c \geq c^*$, the largest component size grows sublinearly with N . This implies the impossibility of an epidemic spreading in the latter case.

Simulation results

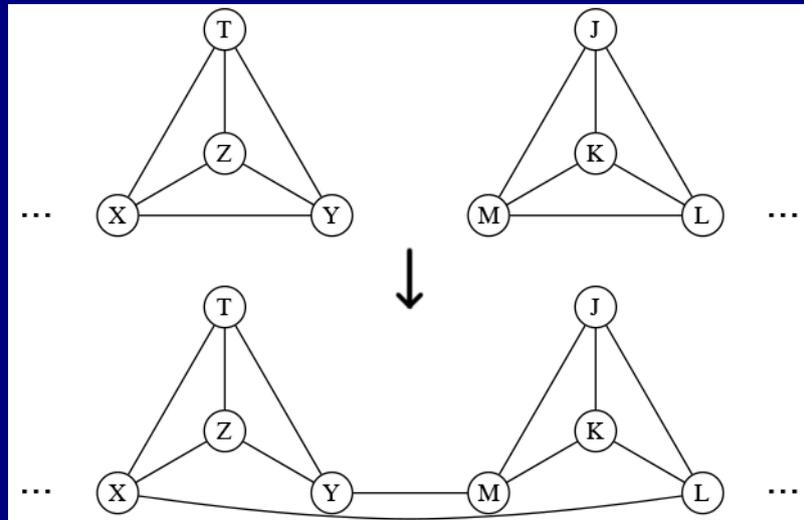
To test the correctness of this result, we simulated ensembles of networks with different k .

We created the networks using exact direct degree-based graph construction and measured the fraction of nodes in the largest component for different values of c .

The results clearly show that s always changes smoothly with c , and that there is no phase transition.



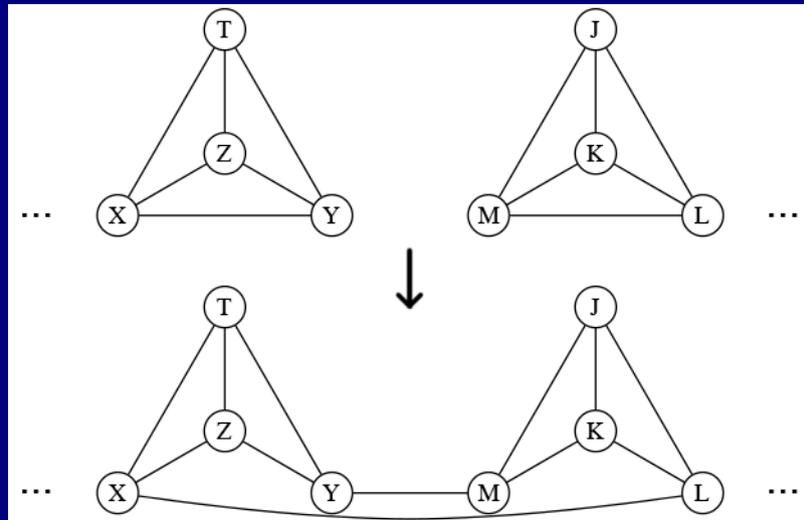
Percolative argument



To understand the absence of a transition, we ask ourselves when a giant component exists, i.e., when the size of the largest component is proportional to the number of nodes N .

For very large N and very small ε , one can generate the ensemble of networks with $c=1-\varepsilon$ starting from $c=1$ and joining cliques with probability ε .

Percolative argument



These new connections establish “external” links σ between clustered neighbourhoods that were formerly isolated.

If the placement of the external links is independent, we can use the Molloy-Reed criterion to determine the existence of a giant component.

Molloy-Reed

Applied to our case, the Molloy-Reed criterion states that a giant component exists if

$$\Sigma \equiv \langle \sigma^2 \rangle - 2 \langle \sigma \rangle > 0.$$

Since ε is small, the probability $P(0)$ of having a neighbourhood with $\sigma=0$ is $O(1)$. Similarly $P(2)=O(\varepsilon)$ and $P(4)=O(\varepsilon^2)$.

But then, $\Sigma=O(\varepsilon^2)>0$. Thus, we expect a giant component to exist for any value of c .

Upper bound

Since we now always expect a giant component, all we need to do is simply finding its actual functional dependence on c .

To do so, notice that the clustering coefficient is the probability that any two nodes which share a common neighbour are linked.

Then, imagine we build a network by randomly linking its nodes, and consider a local neighbourhood.

Upper bound

Under these assumptions, the probability that a node and its k neighbours form a clique of size $k+1$ is

$$p = c^{\frac{k(k-1)}{2}}.$$

Then, to find the total number of cliques, we can multiply this by the number of nodes N , and divide by $k+1$, to account for the fact that in a clique the neighbours of all the nodes are linked amongst each other:

$$n_c = \frac{N}{k+1} c^{\frac{k(k-1)}{2}}.$$

Upper bound

But each clique decreases the fraction of nodes in the largest component by $(k+1)/N$, which means that

$$s(c) = 1 - c^{\frac{k(k-1)}{2}}.$$

Then, the equation above is a simple analytical expression that describes the behaviour of s when c is small enough.

In fact, as we will see, the equation we just derived is an upper bound for the actual relative largest component size.

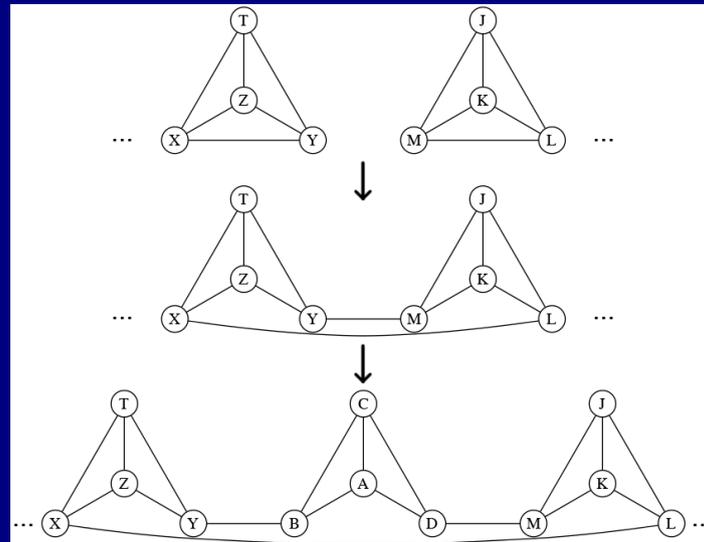
Lower bound

To find a complete description of $s(c)$ we need to use some combinatorial argument. First, it's straightforward to see that

$$\begin{aligned} s(0) &= 1 \\ s(1) &= 0 \\ \left. \frac{ds}{dc} \right|_{c=0^+} &= 0. \end{aligned}$$

Then, we estimate the derivative of s close to $c=1$ with the ratio of the finite differences of s and c .

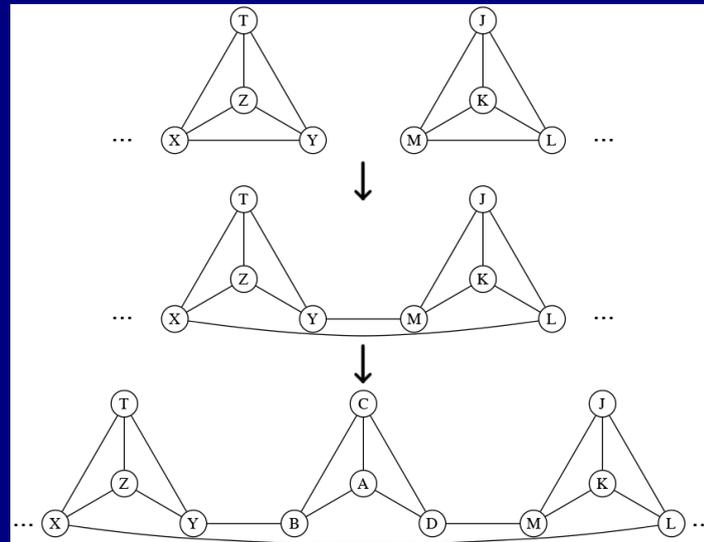
Lower bound



Consider a network with the highest possible value of $c < 1$. We can obtain it by starting from the $c = 1$ case and joining only 2 cliques.

Now, the network with the next highest possible value of $c < 1$ can be obtained by repeating the clique joining procedure.

Lower bound

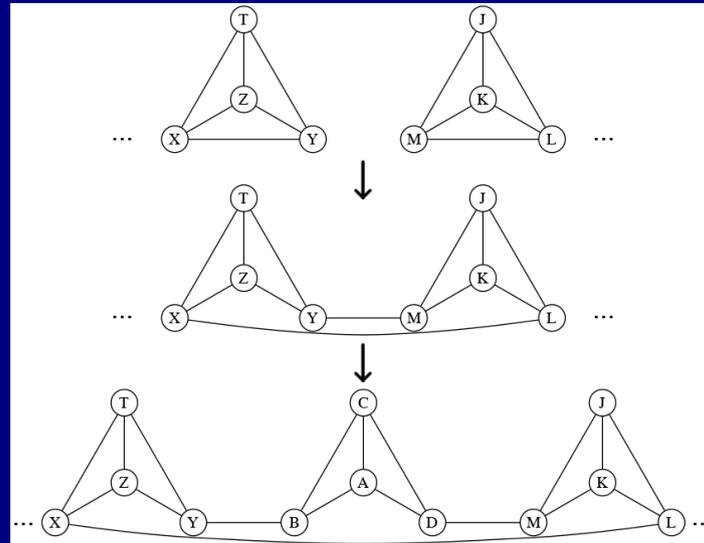


But every time we add a new clique, decreasing c , $k+1$ nodes join the largest component. Thus we can write

$$\Delta S = -\frac{k+1}{N}.$$

Lower bound

$$\Delta s = -\frac{k+1}{N}$$



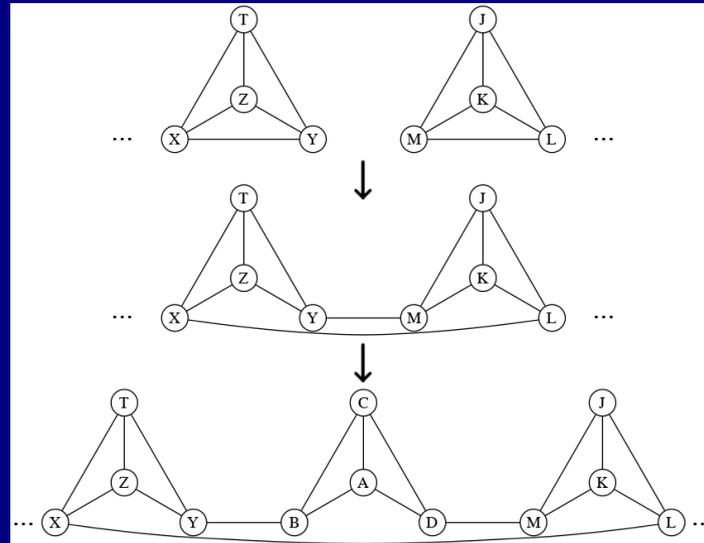
To find Δc , first note that c is the mean of the local clustering coefficients, defined for each node as the number of triangles it belongs to divided by the maximum number of possible links amongst its neighbours:

$$c_L = \frac{2T}{k(k-1)}$$

Lower bound

$$\Delta s = -\frac{k+1}{N}$$

$$c_L = \frac{2T}{k(k-1)}$$



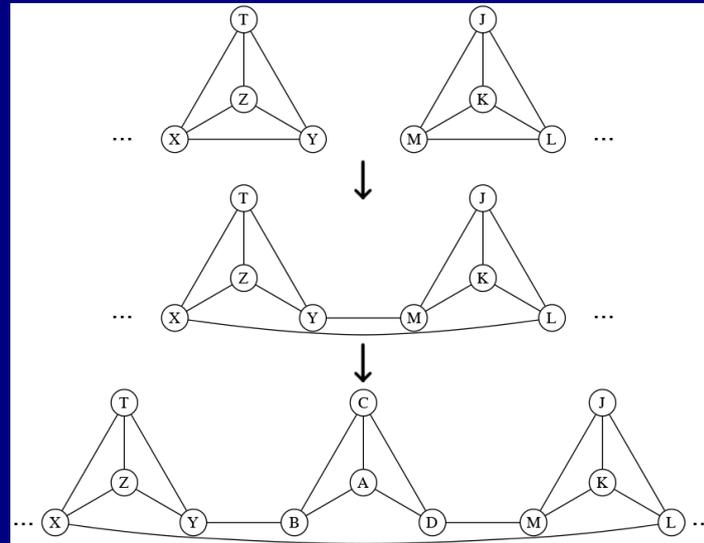
Then, notice that the rewiring breaks 1 triangle for each node that does **not** take part in the rewiring. The rewired nodes, instead, are left with $k-1$ neighbours that form triangles.

So, if we join i cliques, we get $i(k-1)$ nodes with $[k(k-1)/2 - 1]$ triangles each, and $2i$ nodes with $(k-1)(k-2)/2$ triangles each.

Lower bound

$$\Delta s = -\frac{k+1}{N}$$

$$c_L = \frac{2T}{k(k-1)}$$



Thus, the number of triangles in the cliques is

$$T_1 = i(k-1) \left[\frac{k(k-1)}{2} + k - 3 \right]$$

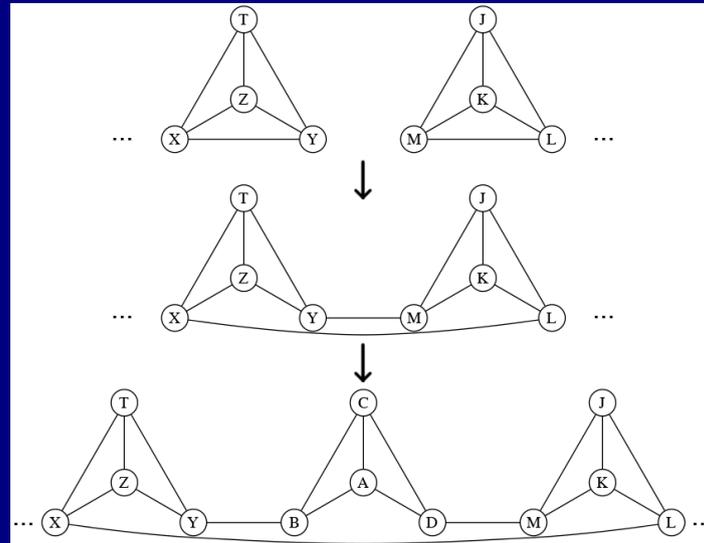
and their contribution is

$$c_1 = \frac{i}{N} \left(k + 1 - \frac{6}{k} \right).$$

Lower bound

$$\Delta s = -\frac{k+1}{N}$$

$$c_L = \frac{2T}{k(k-1)}$$



$$c_1 = \frac{i}{N} \left(k+1 - \frac{6}{k} \right)$$

Finally, there are the other $[N-i(k+1)]$ nodes that did not participate in the rewiring.

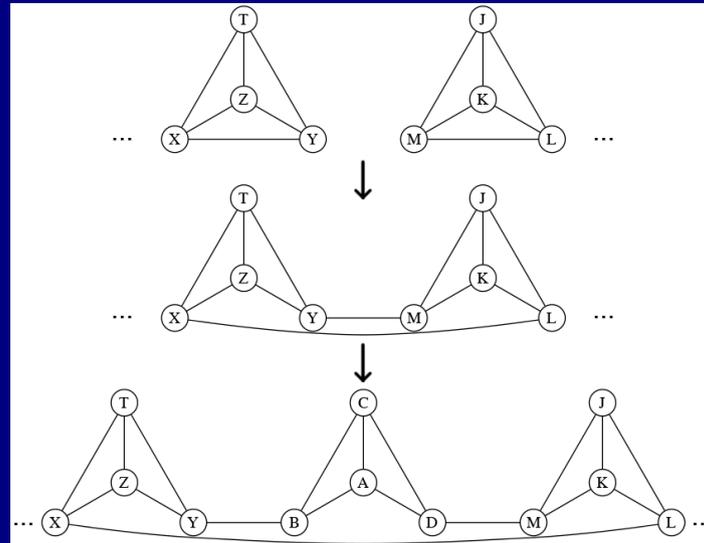
As they are still maximally clustered, their contribution is straightforward to compute and it is simply

$$c_2 = 1 - \frac{i(k+1)}{N}.$$

Lower bound

$$\Delta s = -\frac{k+1}{N}$$

$$c_L = \frac{2T}{k(k-1)}$$



$$c_1 = \frac{i}{N} \left(k+1 - \frac{6}{k} \right)$$

$$c_2 = 1 - \frac{i(k+1)}{N}$$

Summing c_1 and c_2 gives

$$c = 1 - \frac{6i}{kN}$$

This means that every time a clique is added, c decreases of

$$\Delta c = \frac{6}{kN}$$

Lower bound

$$\Delta s = -\frac{k+1}{N}$$

$$\Delta c = \frac{6}{kN}$$

Having computed the finite differences, we can estimate the derivative of s close to $c=1$:

$$\left. \frac{ds}{dc} \right|_{c=1^-} = -\frac{k(k+1)}{6}.$$

Given the previous conditions, this implies that, for high c ,

$$s(c) = 1 - c \frac{k(k+1)}{6}.$$

This is a lower bound for the actual dependence.

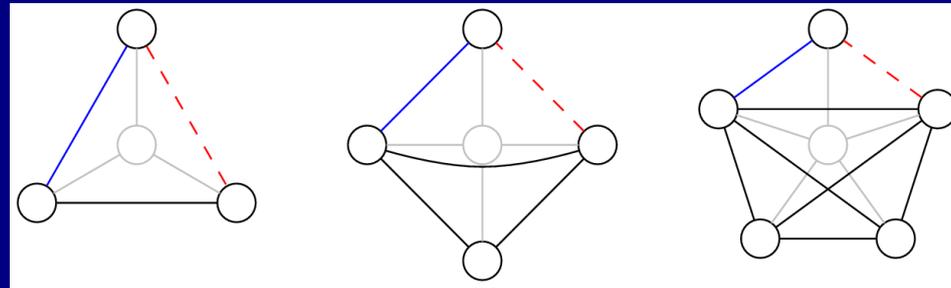
Crossover

What we saw indicates the existence of two regimes, for low and high clustering, with the general form of $s(c)$ given by the crossover behaviour between these two bounds.

To find an explicit form for the full behaviour, first note that every regular network is locally a k -star, that is, a hub node connected to k leaves.

Such a neighbourhood is highly clustered if it has enough links to guarantee that each leaf participates in at least 1 triangle made entirely of leaves.

Crossover



To find how many links are needed, compute how many can be placed while keeping one leaf triangle-free.

This number is found by placing all possible links between $(k-1)$ leaves, and then adding one single further link.

So, for a neighbourhood to be little clustered, the links can be at most $\lfloor (k-1)(k-2)/2 + 1 \rfloor$.

Crossover

But each link between two leaves exists exactly with probability c . Then, the probability that a neighbourhood is highly clustered is simply

$$\begin{aligned} q_{c,k} &\equiv \sum_{x=0}^{\frac{(k-1)(k-2)}{2}+1} \left(1 - \text{Binom} \left(x \mid \frac{k(k-1)}{2}, c \right) \right) \\ &= I_c \left(\frac{1}{2}(k-1)(k-2)+2, k-2 \right). \end{aligned}$$

where

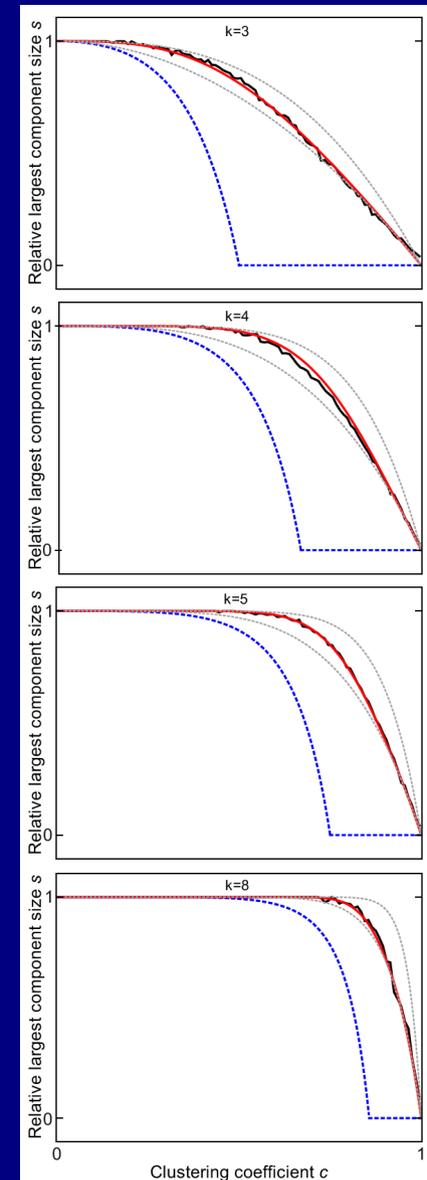
$$I_t(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^t s^{\alpha-1} (1-s)^{\beta-1} ds.$$

Full expression

Because of the average nature of the global clustering coefficient, these calculations allow us to write directly

$$s(c) = 1 - c^{\frac{k(k-1)}{2}} + q_{c,k} \left[c^{\frac{k(k-1)}{2}} - c^{\frac{k(k+1)}{6}} \right].$$

A comparison with the numerical simulation results shows that our analytical solution closely matches the simulated values, confirming the validity of our approach.



Conclusions

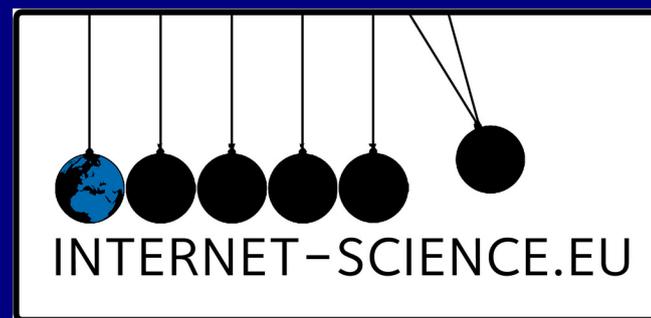
In summary, we have demonstrated that the size of the largest component in regular networks changes smoothly with the clustering c , staying always proportional to the network size.

Thus, regular networks always have a giant component, even for high values of c . This is a new result that is the direct consequence of intrinsic structural properties of the networks, and it is not based on a particular approximation used.

Epidemic spreading is therefore always possible on regular networks, as the maximum fraction of infected individuals does not vanish in the limit of a large population.

Acknowledgements

This work was done in collaboration with Thomas House, of the Warwick Mathematics Institute, the Centre for Complexity Science, and Warwick Infectious Disease Epidemiology Research (WIDER) Centre.



MAX-PLANCK-GESELLSCHAFT



Thank you!